

Coordinated Hybrid Automatic Repeat Request; Extended Version

Behrooz Makki, Tommy Svensson, Thomas Eriksson and Mohamed-Slim Alouini, *Fellow, IEEE*

Abstract—We develop a coordinated hybrid automatic repeat request (HARQ) approach. With the proposed scheme, if a user message is correctly decoded in the first HARQ rounds, its spectrum is allocated to other users, to improve the network outage probability and the users' fairness. The results, which are obtained for single- and multiple-antenna setups, demonstrate the efficiency of the proposed approach in different conditions. For instance, with a maximum of M retransmissions and single transmit/receive antennas, the diversity gain of a user increases from M to $(J+1)(M-1)+1$ where J is the number of users helping that user.

I. INTRODUCTION

Hybrid automatic repeat request (HARQ) is a well-established approach for reliable wireless communication [1]. There are many works improving the performance of HARQ protocols via optimal rate/power allocation, e.g., [1]–[4]. On the other hand, the long-term evolution (LTE) standards provide the capability for dynamic resource allocation in the frequency domain [5, Subsection 16.5.8]. Thus, it is interesting to analyze the performance of HARQ protocols using dynamic frequency allocation.

This letter introduces a *coordinated* HARQ approach. Here, the frequency resources are dynamically allocated among the users based on the HARQ feedback signals. The results are obtained for the repetition time diversity (RTD) and the incremental redundancy (INR) HARQ protocols utilizing single or multiple transmit/receive antennas. As demonstrated in the paper, the advantages of the proposed scheme are: 1) all users benefit from a substantial outage probability improvement and 2) the users' fairness is improved considerably. This is of interest because the fairness has been investigated only in a few HARQ-based systems, e.g., [6], [7]. Moreover, 3) the proposed coordinated approach is useful for buffer-limited transmitters. In harmony with all fairness-based schemes, the coordination may reduce the throughput of the users with the best average channel characteristics slightly. However, the throughput degradation is very limited, as seen in the sequel. Finally, the coordination scales up the diversity gain of the users substantially.

The problem setup of the paper is different from [1]–[3] (resp. [4]) that optimize the performance of single-user (resp. cognitive radio) systems via rate/power adaptation in power-limited (resp. interference-limited) conditions. Also, we investigate a different problem from [6] (resp. [7]) which analyzes the fairness-adaptive throughput optimization in HARQ-based systems using adaptive modulation (resp. the fairness in relay-HARQ systems). Finally,

our discussions on the users' message decoding probabilities, the diversity gain and the fairness have not been presented before¹.

II. SYSTEM MODEL

Consider an HARQ protocol with a maximum of M retransmissions. Also, define a packet as the transmission of a codeword along with all its possible retransmissions and let P be the transmission power for each frequency band. We study block-fading conditions where the channel coefficients remain constant in each retransmission and then change to other values based on their probability density functions (pdf:s). At time slot t , the channel coefficient associated with the i th frequency band is represented by ${}_i h(t)$ and we define ${}_i g(t) \doteq |{}_i h(t)|^2$ which is referred to as the channel gain. For Rayleigh-fading channels, on which we focus, the channel gains follow the pdfs $f_{i,g}(x) = {}_i \lambda e^{-{}_i \lambda x}$, $x \geq 0$, where ${}_i \lambda$ is the fading parameter of the i th channel. In each link, the channel coefficient is assumed to be known by the receiver. However, there is no instantaneous channel state information available at the transmitters.

Coordination Model: The transmitter has access to K frequency bands each having normalized bandwidth $W = 1$ (it is straightforward to extend the results to the cases with different bandwidths). The data transmission protocol is designed as follows. With K users, separate frequency bands are first allocated to the users. If none of the users can correctly decode their corresponding codewords (resp. all users correctly decode their corresponding messages), there is no coordination between the frequency resources, and each user receives its corresponding retransmission (resp. a new message transmission) in the next time slot. The frequency coordination occurs if some of the users successfully decode their corresponding codewords, while the other users cannot. In this case, all frequency resources of the next slot are allocated to the users with unsuccessful message decoding, for which the messages are retransmitted. Denoting the complement of the event s by \bar{s} and $A_n B_m$ as the event that users A and B correctly decode their corresponding messages in rounds n and m , respectively, the following example demonstrates the data transmission protocol for the simplest case with $K = 2$ and $M = 2$ (Also, an illustrative example of the cooperation approach is given in Fig.3 at the end of the paper).

Example: Start the data transmission by sending separate messages to users A and B. The following cases may occur in the next time slot:

- If both users correctly decode their corresponding messages, represented by the event $A_1 B_1$, a new packet transmission starts for each user, within its associated frequency band.

Behrooz Makki, Tommy Svensson and Thomas Eriksson are with Chalmers University of Technology, Email: {behrooz.makki, tommy.svensson, thomase}@chalmers.se. Mohamed-Slim Alouini is with the King Abdullah University of Science and Technology (KAUST), Email: slim.alouini@kaust.edu.sa

¹The uploaded file is an extended version of [8].

- If none of the users decode its corresponding message, shown by the event $\bar{A}_1\bar{B}_1$, the data is retransmitted for each user, within its associated frequency band.
- If user A correctly decodes its message while user B cannot, represented by $A_1\bar{B}_1$, both frequency bands of the next slot are allocated to user B. That is, in the next time slot the codeword of user B is retransmitted in two frequency bands, which is the same as two *simultaneous* retransmissions. Finally, the same procedure is considered if user B successfully decodes the message in round 1, while user A cannot.

III. ANALYSIS

In this section, we analyze the users' outage probability and the system throughput. For simplicity, we first concentrate on the special case of $M = K = 2$ with single-antenna transmission. Later, the results are extended to the cases with $M \geq 2$, $K \geq 2$ and multiple-input-multiple-output (MIMO) transmission. We study the system performance for the RTD and the INR HARQ protocols as two efficient schemes leading to high throughput and low outage probability [1]–[4]. Straightforward modifications can be applied for the cases with basic HARQ.

A. RTD Protocol

Using RTD with codewords of length L , Q_A and Q_B information nats are encoded in each codeword of users A and B, respectively. Thus, the initial transmission rates are $R_A \doteq \frac{Q_A}{L}$, $R_B \doteq \frac{Q_B}{L}$. For each user, the same codeword is retransmitted in the successive retransmission rounds and the receiver performs maximum ratio combining (MRC) of all received signals [1]. Hence, the equivalent transmission rates after m retransmissions are $R_{(m),A} = \frac{Q_A}{mL} = \frac{R_A}{m}$ and $R_{(m),B} = \frac{Q_B}{mL} = \frac{R_B}{m}$. Utilizing the first frequency band, the data transmission of user A stops at the end of the first round if $\log(1 + {}_1g(t)P) > R_A$, otherwise the codeword is retransmitted. Thus, with $M = 2$, different events may occur in each time slot, whose probabilities are given by

$$\begin{aligned}
 & \Pr(A_2B_2) + \Pr(\bar{A}_2B_2) + \Pr(A_2\bar{B}_2) + \Pr(\bar{A}_2\bar{B}_2) = \alpha\beta\gamma, \\
 & \Pr(A_2B_1) + \Pr(\bar{A}_2B_1) = \Pr(\bar{A}_1B_1) = \alpha(1 - \beta)\gamma, \\
 & \Pr(A_1B_2) + \Pr(A_1\bar{B}_2) = \Pr(A_1\bar{B}_1) = (1 - \alpha)\beta\gamma, \\
 & \Pr(A_1B_1) = (1 - \alpha)(1 - \beta)\gamma, \\
 & \gamma \doteq \Pr(A_1B_1) + \Pr(\bar{A}_1B_1) + \Pr(A_1\bar{B}_1) + \Pr(\bar{A}_1\bar{B}_1), \\
 & \alpha \doteq \Pr(\log(1 + {}_1g(t)P) < R_A), \\
 & \beta \doteq \Pr(\log(1 + {}_2g(t)P) < R_B).
 \end{aligned} \tag{1}$$

Setting the sum of all possible probabilities equal to 1, the sum probability of all possible events in the first slot of the new packet transmissions, i.e., γ in (1), is found as

$$\gamma = \frac{1}{1 + \alpha + \beta - \alpha\beta}, \tag{2}$$

from which the probabilities $\Pr(A_1B_1)$, $\Pr(A_1\bar{B}_1)$, $\Pr(\bar{A}_1\bar{B}_1)$ and $\Pr(\bar{A}_1B_1)$ are obtained (see (1)).

Given that user A successfully decodes its corresponding message at the end of round 1 while user B cannot decode its associated codeword, two copies of the user B's codeword are retransmitted in the two frequency bands of the next slot. The receiver of user B performs MRC of the three received signals (1 transmission plus 2 retransmissions). Hence, we have

$$\begin{aligned}
 \Pr(A_1\bar{B}_2) &= \gamma \Pr\left(\log(1 + {}_1g(t)P) \geq R_A \cap \right. \\
 & \left. \log(1 + ({}_2g(t) + {}_2g(t+1) + {}_1g(t+1))P) < R_B\right), \tag{3}
 \end{aligned}$$

and $\Pr(A_1B_2) = (1 - \alpha)\beta\gamma - \Pr(A_1\bar{B}_2)$. Here, we have used the fact that with the signal-to-noise ratio (SNR) of SNR_i for the i th received signal the maximum decodable rate is

$$U_{(m)}^{\text{RTD}} = \frac{1}{m} \log\left(1 + \sum_{i=1}^m \text{SNR}_i\right), \tag{4}$$

if the same codeword is retransmitted m times [1, Section III].

For Rayleigh-fading channels, (3) is found as²

$$\begin{aligned}
 \Pr(A_1\bar{B}_2) &= \gamma(1 - \alpha)\Phi, \quad \alpha = 1 - e^{-1^{\lambda C_A}}, \\
 \Phi &= \Pr({}_2g(t) + {}_2g(t+1) + {}_1g(t+1) < C_B) \\
 &= \int_0^{C_B} \int_0^{C_B-x} f_{1g}(x) f_{2g}(y) \Pr({}_2g(t+1) < C_B - x - y) dx dy \\
 &= \int_0^{C_B} \int_0^{C_B-x} {}_1\lambda e^{-1^{\lambda x}} {}_2\lambda e^{-2^{\lambda y}} (1 - e^{-2^{\lambda(C_B-x-y)}}) dx dy \\
 &= 1 - e^{-1^{\lambda C_B}} + \frac{e^{-2^{\lambda C_B}} - e^{-1^{\lambda C_B}}}{\frac{2^{\lambda}}{1^{\lambda}} - 1} + \frac{C_B e^{-2^{\lambda C_B}}}{\frac{1}{1^{\lambda}} - \frac{1}{2^{\lambda}}} \\
 &+ \frac{{}_1\lambda {}_2\lambda}{(1^{\lambda} - 2^{\lambda})^2} (e^{-2^{\lambda C_B}} - e^{-1^{\lambda C_B}}), \tag{5}
 \end{aligned}$$

where $C_A \doteq \frac{e^{R_A}-1}{P}$, $C_B \doteq \frac{e^{R_B}-1}{P}$. The other probabilities, e.g., $\Pr(\bar{A}_2B_1)$, $\Pr(\bar{A}_2\bar{B}_1)$, $\Pr(A_1B_2)$ and the outage probabilities, e.g., $\Pr(\text{Outage}_B) = \Pr(A_2\bar{B}_2) + \Pr(\bar{A}_2\bar{B}_2) + \Pr(A_1\bar{B}_2)$ are found with the same procedure as in (5) leading to

$$\begin{aligned}
 \Pr(\text{Outage}_B) &= \gamma\alpha(1 - e^{-2^{\lambda C_B}} - {}_2\lambda C_B e^{-2^{\lambda C_B}}) \\
 &+ \gamma(1 - \alpha)\left(1 - e^{-1^{\lambda C_B}} + \frac{e^{-2^{\lambda C_B}} - e^{-1^{\lambda C_B}}}{\frac{2^{\lambda}}{1^{\lambda}} - 1} + \frac{C_B e^{-2^{\lambda C_B}}}{\frac{1}{1^{\lambda}} - \frac{1}{2^{\lambda}}}\right. \\
 &+ \left. \frac{{}_1\lambda {}_2\lambda}{(1^{\lambda} - 2^{\lambda})^2} (e^{-2^{\lambda C_B}} - e^{-1^{\lambda C_B}})\right). \tag{6}
 \end{aligned}$$

The throughput (in nats-per-channel-use (npcu)) is defined as

$$\eta = \lim_{N \rightarrow \infty} \frac{\sum_{t=1}^N \tilde{Q}_A(t) + \sum_{t=1}^N \tilde{Q}_B(t)}{NL}, \tag{7}$$

where $\sum_{t=1}^N \tilde{Q}_A(t)$ and $\sum_{t=1}^N \tilde{Q}_B(t)$ denote the total number of information nats that are successfully decoded by users A and B, respectively, in N time slots [1]. Using (1)–(5), the law of large numbers and $N \rightarrow \infty$ time slots, the total number of information nats successfully decoded by user A is found as

$$\begin{aligned}
 \sum_{t=1}^N \tilde{Q}_A(t) &= Q_A N \left(\Pr(A_1\bar{B}_1) + \Pr(A_1B_1) \right. \\
 &+ \left. \Pr(A_2B_1) + \Pr(A_2B_2) + \Pr(A_2\bar{B}_2) \right).
 \end{aligned}$$

Thus, from $R_A = \frac{Q_A}{L}$, $R_B = \frac{Q_B}{L}$, the throughput is obtained by

$$\begin{aligned}
 \eta &= R_A \left(\Pr(A_1\bar{B}_1) + \Pr(A_1B_1) + \Pr(A_2B_1) + \Pr(A_2B_2) \right. \\
 &+ \left. \Pr(A_2\bar{B}_2) \right) + R_B \left(\Pr(\bar{A}_1B_1) + \Pr(A_1B_1) + \Pr(A_1B_2) \right. \\
 &+ \left. \Pr(A_2B_2) + \Pr(\bar{A}_2B_2) \right). \tag{8}
 \end{aligned}$$

²The analytical results are given for ${}_i\lambda \neq {}_j\lambda, i \neq j$. Straightforward modifications should be applied for the cases with ${}_i\lambda = {}_j\lambda, i \neq j$.

B. INR Protocol

Using INR, new codewords are sent in the successive retransmission rounds and the message is decoded by the receivers using all previously received signals of the packet. In this case, the results of [1]–[4] can be used to rephrase the INR-based probability terms as, e.g.,

$$\begin{aligned} \Pr(A_1 B_2) &= \gamma \Pr \left(\log(1 + {}_1g(t)P) \geq R_A \cap \right. \\ &\log(1 + {}_2g(t)P) < R_B \leq \log(1 + {}_2g(t)P) \\ &\left. + \log(1 + {}_2g(t+1)P) + \log(1 + {}_1g(t+1)P) \right). \end{aligned} \quad (9)$$

That is, the achievable rate terms $U_{(m)}^{\text{RTD}} = \frac{1}{m} \log(1 + \sum_{i=1}^m \text{SNR}_i)$ of the RTD, i.e., (4), are replaced by the terms

$$U_{(m)}^{\text{INR}} = \frac{1}{m} \sum_{i=1}^m \log(1 + \text{SNR}_i) \quad (10)$$

in the INR, and the probabilities are recalculated. This is the only modification required for the INR, compared to the RTD, and the rest of the discussions remain the same as before.

C. Extension of Results to Arbitrary Number of Retransmissions

The results can be extended to the case with a maximum of $M \geq 2$ retransmissions. Here, the probability that, for instance, users A and B successfully decode their messages at the n th and m th, $n \leq m$, rounds of the RTD, respectively, is obtained by

$$\begin{aligned} \Pr(A_n B_m) &= \gamma \Pr \left(\log(1 + P \sum_{i=0}^{n-2} {}_1g(t+i)) < \right. \\ R_A &\leq \log(1 + P \sum_{i=0}^{n-1} {}_1g(t+i)) \cap \\ \log(1 + P \sum_{i=0}^{m-2} {}_2g(t+i) + P \sum_{i=n}^{m-2} {}_1g(t+i)) &< R_B \leq \\ \log(1 + P \sum_{i=0}^{m-1} {}_2g(t+i) + P \sum_{i=n}^{m-1} {}_1g(t+i)) &\left. \right), n \leq m, \end{aligned} \quad (11)$$

and the other terms, e.g., η , $\Pr(\text{Outage}_A)$ and $\Pr(\text{Outage}_B)$ are rephrased correspondingly.

In (5), we presented a closed-form expression for the probabilities, e.g., $\Pr(A_1 \bar{B}_2)$, with $M = 2$. Theorem 1 extends the results to the cases with arbitrary number of retransmissions.

Theorem 1: For Rayleigh fading channels, the throughput and the outage probability of the proposed RTD- and INR-based schemes are obtained via the following equalities, respectively

$$\begin{aligned} \Pr(\log(1 + P(\sum_{k=0}^{n-1} {}_1g(t+k) + \sum_{l=t'}^{t'+m-1} {}_2g(t+l))) < x) \\ &= \mathcal{W}(e^x - 1) - \mathcal{W}(0), \\ \mathcal{W}(x) &\doteq - \sum_{k=1}^n \frac{a_k \Gamma(k, \lambda x)}{(k-1)!} - \sum_{k=1}^m \frac{b_k \Gamma(k, \lambda x)}{(k-1)!}, \\ a_k &\doteq (-\frac{1}{2\lambda})^{n-k} \binom{n+m-k-1}{n-k} (1 - \frac{1}{2\lambda})^{-(n+m-k)}, \\ b_k &\doteq (-\frac{2}{1\lambda})^{m-k} \binom{n+m-k-1}{m-k} (1 - \frac{2}{1\lambda})^{-(m+n-k)}, \end{aligned} \quad (12)$$

$$\begin{aligned} \Pr(\sum_{k=0}^{n-1} \log(1 + {}_1g(t+k)P) + \sum_{l=t'}^{t'+m-1} \log(1 + {}_2g(t+l)P) < x) \\ &= 1 - e^{-\frac{\lambda n + 2\lambda m}{P}} \times \\ &\mathcal{Y}_{n+m+1,1}^{n+m+1,0} \left[\frac{1}{P^{n+m}} e^x \middle| \underbrace{(1,1,0), (0,1,0), (1,1,\frac{1}{P}), \dots, (1,1,\frac{1}{P})}_{n \text{ times}}, \dots, \right. \\ &\left. \underbrace{\dots, (1,1,\frac{2}{P}), \dots, (1,1,\frac{2}{P})}_{m \text{ times}} \right] \forall t', n, m. \end{aligned} \quad (13)$$

Here, $\Gamma(\cdot, \cdot)$ and $\mathcal{Y}_{s_1, s_2}^{s_3, s_4}[\cdot]$ are the incomplete Gamma function and the generalized upper incomplete Fox'H function [9], respectively. Also, $\binom{n}{k}$ denotes the “ n choose k ” operator.

Proof: Following (3)–(11) and the same discussions as in [1, Section IV], it is straightforward to show that the throughput and the outage probability of the RTD- and INR-based schemes can be represented as monotonic functions of the probabilities

$$\begin{cases} \pi_{\text{RTD}} = \Pr \left(\log(1 + P(\sum_{k=0}^{n-1} {}_1g(t+k) + \sum_{l=t'}^{t'+m-1} {}_2g(t+l))) < x \right), & \text{For RTD} \\ \pi_{\text{INR}} = \Pr \left(\sum_{k=0}^{n-1} \log(1 + {}_1g(t+k)P) + \sum_{l=t'}^{t'+m-1} \log(1 + {}_2g(t+l)P) < x \right), & \text{For INR.} \end{cases}$$

To find π_{RTD} (and then the throughput and outage probability), we use Laplace transform $\mathcal{L}\{\cdot\}$ and its inverse $\mathcal{L}^{-1}\{\cdot\}$ to write

$$\begin{aligned} \pi_{\text{RTD}} &\stackrel{(a)}{=} \int_0^{e^x-1} \mathcal{L}^{-1} \left\{ \frac{1}{(1 + \frac{Ps}{\lambda})^n (1 + \frac{Ps}{2\lambda})^m} \right\} dz \\ &\stackrel{(b)}{=} \int_0^{e^x-1} \mathcal{L}^{-1} \left\{ \sum_{k=1}^n \frac{a_k}{(1 + \frac{Ps}{\lambda})^k} + \sum_{k=1}^m \frac{b_k}{(1 + \frac{Ps}{2\lambda})^k} \right\} dz \\ &\stackrel{(c)}{=} \int_0^{e^x-1} \left(\sum_{k=1}^n \frac{a_k \lambda^k z^{k-1} e^{-\frac{1}{P}z}}{(k-1)!} + \sum_{k=1}^m \frac{b_k 2^k z^{k-1} e^{-\frac{1}{P}z}}{(k-1)!} \right) dz \\ &= \mathcal{W}(e^x - 1) - \mathcal{W}(0). \end{aligned}$$

Here, (a) follows from the fact that the pdf of the sum of independent random variables is obtained by the convolution of their pdfs and $\mathcal{L}\{f_i \theta\} = (1 + \frac{Ps}{\lambda})^{-1} \cdot \theta \doteq {}_1gP$. Then, (b) is obtained by partial fraction of $(1 + \frac{Ps}{\lambda})^{-n} (1 + \frac{Ps}{2\lambda})^{-m}$, with fraction coefficients a_k, b_k given in (12). Also, (c) is derived by inverse Laplace transform and some manipulations.

Finally, the probabilities π_{INR} of the INR are obtained by appropriate parameter setting in [9, eq. 18] leading to (13). Then, having $\pi_{\text{RTD}}, \pi_{\text{INR}}$, the system performance is analyzed with the same method as in (1)–(11). ■

Note that the generalized upper incomplete Fox'H function has an efficient MATHEMATICA implementation [9].

D. Multiple-Antenna Scenario

From another perspective, we can extend the results to the case with MIMO transmission; consider a setup with u transmit antennas and v receive antennas for each user. Let $\mathbf{H}_i(t) \in \mathcal{C}^{v \times u}$ denote the complex channel matrix associated with the i th frequency band at time slot t . Also, represent the $u \times u$ identity matrix by \mathbf{I}_u . Using isotropic input distribution over all transmit antennas, the same procedure as in [2, Section III.C] can be used to rephrase the achievable rate term of the RTD, i.e., (4), as

$$U_{(n,m),B}^{\text{RTD}} = \frac{1}{2m-n} \log \det(I_{(2m-n)v} + \frac{P}{u} H_{(n,m),B}^{\text{RTD}} (H_{(n,m),B}^{\text{RTD}})^*),$$

$$H_{(n,m),B}^{\text{RTD}} = [{}_2H^T(t) \dots {}_2H^T(t+m-1) {}_1H^T(t+n) \dots {}_1H^T(t+m-1)]^T,$$

if the data retransmissions of users A and B continue up to the end of rounds n and m , $n \leq m$, respectively. Here, $\det(X)$, X^T and X^* represent the determinant, the transpose and the Hermitian of the matrix X , respectively. For INR, we have

$$U_{(n,m),B}^{\text{INR}} = \frac{1}{2m-n} \left(\sum_{i=0}^{m-1} \log \det(I_v + \frac{P}{u} {}_2H(t+i) {}_2H(t+i)^*) \right. \\ \left. + \sum_{i=n}^{m-1} \log \det(I_v + \frac{P}{u} {}_1H(t+i) {}_1H(t+i)^*) \right). \quad (14)$$

In this way, using

$$U_{(n,0),A}^{\text{RTD}} \doteq \frac{1}{n} \log \det \left(I_{nv} + \frac{P}{u} H_{(n,0),A}^{\text{RTD}} (H_{(n,0),A}^{\text{RTD}})^* \right),$$

$$H_{(n,0),A}^{\text{RTD}} = [{}_1H^T(t) \dots {}_1H^T(t+n-1)]^T, \quad (15)$$

the probabilities, e.g., (11), are obtained by

$$\Pr(A_n B_m) = \Pr \left(U_{(n-1,0),A}^{\text{RTD}} < \frac{R_A}{n-1} \cap \right. \\ \left. U_{(n,0),A}^{\text{RTD}} \geq \frac{R_A}{n} \cap U_{(n,m-1),B}^{\text{RTD}} < \frac{R_B}{2(m-1)-n} \cap \right. \\ \left. U_{(n,m),B}^{\text{RTD}} \geq \frac{R_B}{2m-n} \right), \quad (16)$$

for RTD, while the rest of the arguments remain the same as in Subsection III.A. Finally, we can use (14) and the same procedure as in (15)-(16) to derive the probabilities for the INR.

E. Coordination with $K > 2$ Users

The system performance in the presence of $K > 2$ users depends on the designed coordination rules. However, Theorem 2 shows that assigning the free frequency bands of J users to a user scales up its diversity gain, i.e., the negative of the slope of its outage probability curve at high SNRs, to $d = (J+1)(M-1)+1$.

Theorem 2: Using INR, the diversity gain of a user is $d = (J+1)(M-1)+1$, if the coordination rule can provide it with the free frequency bands of J users.

Proof: With no loss of generality, let us consider the K th (the last) user and assume that it can utilize the free frequency resources of the first J users. The diversity gain $d_K = -\lim_{P \rightarrow \infty} \frac{\log(\Pr(\text{Outage}_K))}{\log P}$ [3, eq. 14] of user K is found as

$$d_K = -\lim_{P \rightarrow \infty} \frac{\log \left(\Pr \left(\bigcup_{n_j=1, \dots, M, j=1, \dots, J} \xi_K(n_1, \dots, n_J) \right) \right)}{\log P}$$

$$\stackrel{(a)}{=} -\lim_{P \rightarrow \infty} \frac{\log \left(\Pr \left(\left(\bigcap_{j=1, \dots, J} \omega_j(n_j) \right) \cap \phi_K(n_1, \dots, n_J) \right) \right)}{\log P}$$

$$\stackrel{(b)}{=} -\lim_{P \rightarrow \infty} \frac{\log \left(\prod_{j=1}^J (P^{-(n_j-1)} - P^{-n_j}) P^{-(M+\sum_{j=1}^J (M-n_j))} \right)}{\log P}$$

$$= -\lim_{P \rightarrow \infty} \frac{\log(P^{-(\sum_{j=1}^J (n_j-1) + M + \sum_{j=1}^J (M-n_j))})}{\log P}$$

$$= (J+1)(M-1)+1,$$

$$\xi_K(n_1, \dots, n_J) \doteq \{\text{Outage}_K \& c_j = n_j\},$$

$$\omega_j(n_j) \doteq \left\{ \sum_{t=1}^{n_j-1} \log(1 + {}_jg(t)P) < R \leq \sum_{t=1}^{n_j} \log(1 + {}_jg(t)P) \right\}$$

$$\phi_K(n_1, \dots, n_J) \doteq \left\{ \sum_{t=1}^M \log(1 + {}_Kg(t)P) + \sum_{j=1}^J \sum_{t=n_j+1}^M \log(1 + {}_jg(t)P) < R \right\}.$$

Here, c_j is the indicator of the slot number in which the j th user message is decoded. Then, $\xi_K(n_1, \dots, n_J)$ is the event of successful decoding for users $j = 1, \dots, J$ at slots n_1, \dots, n_J and outage for user K . Also, $\omega_j(n_j)$ is the event of successful decoding for the j th user in the n_j th round and $\phi_K(n_1, \dots, n_J)$ is the K th user outage event while utilizing the $j = 1, \dots, J$ users' frequency bands in rounds $n_j + 1, \dots, M, j = 1, \dots, J$. Then, (a) is based on the fact that 1) $\xi_K(n_1, \dots, n_J)$'s are disjoint events for different n_j 's, $j = 1, \dots, J$, 2) different terms of the union are of the same order of P and, thus, 3) at $P \rightarrow \infty$ the diversity gain is obtained by considering only one term of the union. Finally, (b) follows from (10) at $P \rightarrow \infty$. ■

Intuitively, the theorem indicates that, at high SNR, e.g., the first J users decode their corresponding messages at their first round, with very high probability. Thus, e.g., the K th user can utilize its own M retransmissions and the remaining $J(M-1)$ retransmissions of users $j = 1, \dots, J$. Consequently, we have $d_K = (J+1)(M-1)+1$. Then, the diversity gain of the whole system containing K users is given by $\min_{k=1, \dots, K} \{d_k\}$. As an example, with $K = 2$ users the diversity gain of the coordinated scheme is increased to $2M-1$, compared to the non-coordinated setting for which we have $d_{\text{non-coordinated}} = M$, independently of the number of users. Note that the theorem is presented for the single-antenna setting, while it can be extended for the MIMO setup. Moreover, although the theorem is proved for the INR scheme, the same point holds for the RTD as well (also, see Fig. 2 for examples). Finally, the performance gain is at the cost of coordination overhead mainly at the receivers receiving messages in different frequency slots.

IV. RESULTS AND CONCLUSIONS

The simulation results of Fig.1 are obtained for $K = 2$ users. Here, except for the MIMO setup where the probabilities, e.g., (16) are calculated numerically, the results are obtained both analytically and via Monte Carlo simulations which lead to the same results. Therefore, to avoid too much information in each figure, we plot only one of them. Using the INR, Fig.1a compares the users' outage probability in different schemes. As shown, the coordination decreases the users' outage probability substantially. Also, the impact of coordination on the outage probability increases with the SNR/maximum number of retransmissions M . Finally, as shown in the figure, the negative of the slope of the

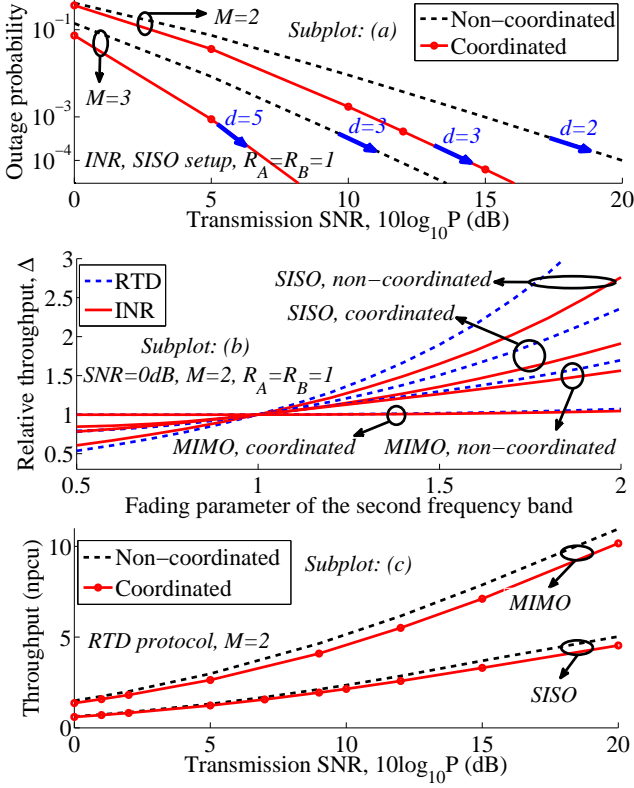


Figure 1. Comparison between the coordinated and non-coordinated schemes from (a): outage probability, (b): fairness and (c): throughput perspectives. $K = 2$, $\lambda_1 = \lambda_2 = 1$ (except in figure (b), which is for different values of λ_2). In figures (a)-(b), $R_A = R_B = 1$. In figure (c), the rates R_A, R_B are optimized in each SNR, to maximize throughput. For the MIMO setup, we set $u = v = 2$.

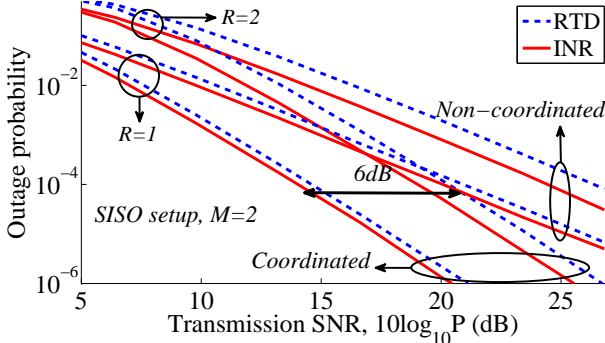


Figure 2. Outage probability in the cases with $K = 3$ users, SISO setup, $M = 2$. The initial rate of all users is set to $R = 1, 2$.

outage probability curves at moderate/high SNRs is the same as the diversity gain derived in Theorem 2. For instance, with $K = 2$ and $M = 2$, the diversity gain of the coordinated and non-coordinated schemes are $d = 3$ and $d = 2$, respectively.

To study the fairness, we plot $\Delta = \frac{\eta_A}{\eta_B}$, i.e., the ratio of the users' throughput, for different values of λ_2 . Moreover, optimizing the transmission rates by exhaustive search, Fig.1c shows the system throughput (8) for various schemes. As it is seen, the proposed coordinated HARQ scheme improves the users' fairness considerably (Fig.1b), and the throughput loss is negligible in the considered range of SNR (Fig.1c). Also, the users' fairness, outage probability and throughput are improved by increasing the number of transmit/receive antennas.

Setting $\lambda_i = 1, \forall i$, Fig.2 studies the outage probability in the cases with $M = 2$ and $K = 3$ users. Here, the initial rate of all users is set to $R = 1, 2$. Also, if only one user cannot decode its message correctly at the end of round 1, it receives

all frequency bands of the second round. Then, in the cases with two unsuccessful users at the end of the first round, the three frequency slots of the second round are randomly allocated to those users such that one of them receives two frequency slots (and the other receives one). As seen, the INR and the RTD schemes have the same diversity gain (see Theorem 2 and its following discussions). Also, the coordination leads to considerable improvements in the energy efficiency. As an example, with $R = 1$ and outage probability 10^{-4} the coordination improves the energy efficiency of the INR approach by 6dB.

To conclude, as demonstrated both theoretically and via simulations, the proposed coordinated HARQ approach leads to considerable users' outage probability and fairness improvement, with limited throughput degradation.

REFERENCES

- [1] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1971–1988, July 2001.
- [2] C. Shen and M. P. Fitz, "Hybrid ARQ in multiple-antenna slow fading channels: Performance limits and optimal linear dispersion code design," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 5863–5883, Sept. 2011.
- [3] H. El-Gamal, G. Caire, and M.-O. Damen, "The MIMO ARQ channel: Diversity-multiplexing-delay tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3601–3621, Aug. 2006.
- [4] B. Makki, A. Graell i Amat, and T. Eriksson, "HARQ feedback in spectrum sharing networks," *IEEE Commun. Lett.*, vol. 16, no. 9, pp. 1337–1340, Sept. 2012.
- [5] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, *3G Evolution: HSPA and LTE mobile broadband*. Academic press, 2008.
- [6] J. J. Escudero-Garzas, B. Devillers, and A. Garcia-Armada, "Fairness-adaptive goodput-based resource allocation in OFDMA downlink with ARQ," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1178–1192, March 2014.
- [7] Y. Kwon and J. W. Jang, "Improving fairness using ARQ messages in LTE mobile multi-hop relay (MMR) networks," in *WCSP*, Nov. 2009, pp. 1–5.
- [8] B. Makki, T. Svensson, T. Eriksson, and M. S. Alouini, "Coordinated hybrid automatic repeat request," *IEEE Commun. Lett.*, vol. 18, no. 11, pp. 1975–1978, Nov. 2014.
- [9] F. Yilmaz and M.-S. Alouini, "Product of shifted exponential variates and outage capacity of multicarrier systems," in *Eur. Wireless*, 2009, pp. 282–286.

